

阿里妈妈展示广告校准技术演进之路



阿里妈妈

阿里妈妈——展示广告算法团队

黄思光 | 2021-04-28

VISIT...

LANZAROTE
Caliente.COM

大纲

- 背景介绍
- 校准算法演进
- 工程实践
- 总结与展望

To Make It Easy

To Do Marketing Anywhere

阿里妈妈依托集团核心的商业数据和超级媒体矩阵，为客户提供全链路的消费者运营解决方案，让商业营销更简单高效

预估技术的挑战

用户行为概率预估技术（点击率，转化率，短视频3s曝光率等）是搜索、推荐以及广告领域的核心技术能力。



点击率预估模型输出的值是真实点击概率吗？

用户的真实点击概率可以被准确预测吗？

如何衡量模型的准确性？AUC？

预估点击率
(PCTR)



真实点击概率
(ACTR)

优化序关系 (AUC)
(✓)

而非大小准确性 (MSE)
(✗)

预估校准在广告中的必要性

校准技术在**出价准确性/稳定性**（包括算法出价oCPX / Auto-Bidding），**竞价公平性**，**计划冷启动**等方面发挥作用。

举例1：CPC出价下4个广告参竞 $eCPM = bid * pCTR * 1000$

广告	出价 (bid)	pCTR	eCPM	Rank
ad1	1.5	0.8%	12	1
ad2	0.6	1.8%	11	2
ad3	1.0	1.0%	10	3
ad4	0.5	1.4%	7	4

ad3 : PCTR **1.0% → 1.3%**

- 绝对值高估30%，AUC不变
- eCPM **10 → 13**，竞得展示机会
- 广告拿量能力变大，平台期望收入降低

举例2：CPC (AD2&AD4) & CPM (AD1&AD3) 广告混竞

广告	出价	pCTR	eCPM	Rank
AD1	\	1.0%	13	1
AD2	1.0	1.1%	11	2
AD3	\	1.2%	9	3
AD4	0.5	1.4%	7	4

AD2 : PCTR **1.1% → 1.2%**

- AD1多扣费，平台收入增加
- PCTR **1.1% → 1.0%**
- AD1少扣费，平台收入折损

多种Bid类型

多种物料

多种预估模型

广告/推荐混排

校准目标与问题拆解

校准目标

使用户行为的预估值尽可能**逼近真实概率值**，最小化MSE（排序水平不降，绝对值变好）

$$f^* = \arg \min_f \int_x (E[Y|X=x] - f(x))^2 dx$$

相关工作

- 预估模型和校准方法结合
 - 将**先验信息**，**数据采样比例**等导致绝对值偏差的因素引入模型中（或结合常用校准方法）
- 后处理方法（Post-processing）
 - 预估模型和校准算法解耦（轻量级的后处理技术，如Binning，Scaling，Binning-scaling方法）
 - 1) 计算**绝对值准确的后验概率**
 - 2) 构建校准函数对预估值与绝对值进行映射

如何获得准确的后验概率？

关键问题（针对后处理校准方法）

- **校准维度选择问题**：特征越相似的PV，模型预估效果（PCTR高/低估）越接近，校准系数计算越准确
- **数据稀疏问题**：数据量越稀疏，后验统计误差越大，校准系数的计算准确度越低

校准评价指标

Predicted click over click (PCOC)

$$PCOC = \frac{\text{average calibrated probability}}{\text{posterior probability}}$$

PCOC的局限性（举例）：

Samples	Calibrated Prob	Posterior Prob	PCOC
10000	0.2	0.4	0.5（低估）
10000	0.8	0.6	1.3（高估）
Total	(0.2+0.8)/2	(0.4+0.6)/2	1.0（正常？）

Calibration-N (Cal-N)

Distribution Approximation: Aggregate samples with similar calibrated probabilities

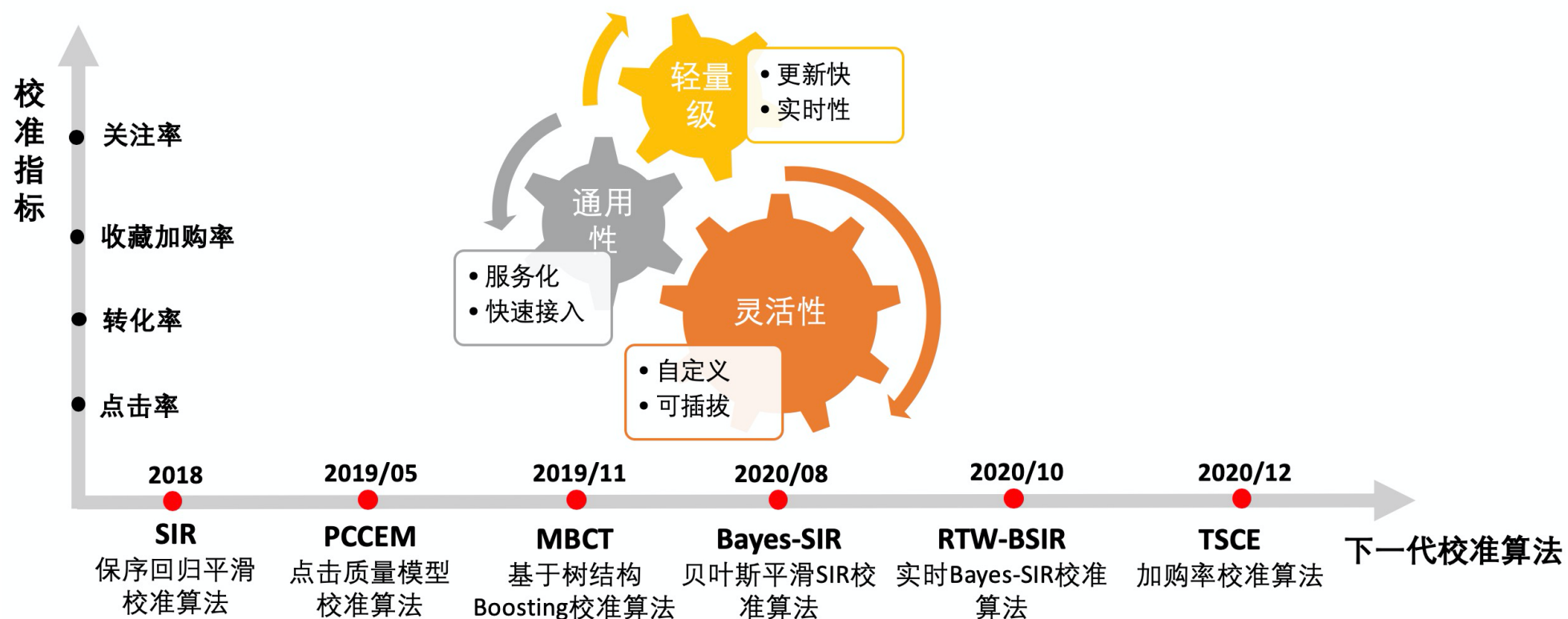
$$\text{Cal-N} = \sqrt{\frac{\sum_{i=1}^N \text{error}_i^2}{N}}, \text{ where } \text{error}_i = \begin{cases} \text{PCOC}_i - 1 & \text{PCOC}_i \geq 1 \\ \frac{1}{\text{PCOC}_i} - 1 & \text{PCOC}_i < 1 \end{cases}$$

Grouped Calibration-N (GC-N)

Domain-specific: weighted average Cal-N of m campaigns

$$\text{GC-N} = \frac{\sum_{j=1}^m w_j \text{Cal-N}_j}{\sum_{j=1}^m w_j}$$

阿里妈妈校准校准算法演进



SIR : 保序回归平滑校准算法 (Smoothed Isotonic Regression)

分桶并计算桶内平均CTR及PCTR

$$CTR_i = \frac{\sum click}{\sum pv} \quad PCTR_i = \frac{\sum pctr}{\sum pv}$$

保序回归(PAV算法)

如果 $CTR_i > CTR_{i+1}$ (违反桶间大小顺序), 则合并桶.

$$CTR_{new} = \frac{pv_i \times CTR_i + pv_{i+1} \times CTR_{i+1}}{pv_i + pv_{i+1}}$$

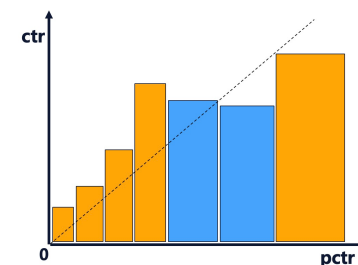
$$PCTR_{new} = \frac{pv_i \times PCTR_i + pv_{i+1} \times PCTR_{i+1}}{pv_i + pv_{i+1}}$$

桶间插值产生分段校准函数

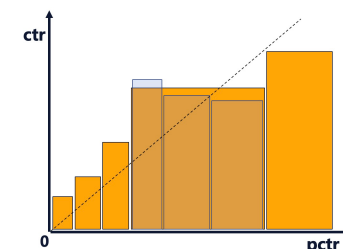
$$a_i = \frac{CTR_{i+1} - CTR_i}{PCTR_{i+1} - PCTR_i} \quad b_i = CTR_i - a_i PCTR_i$$

$$f(x) = a_i x + b_i \quad (if \quad l_i < x \leq u_i)$$

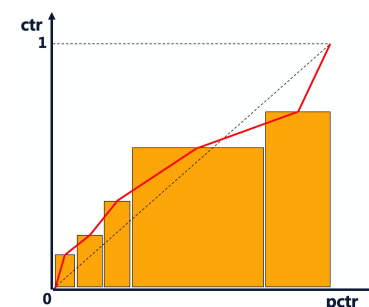
1. 分桶 (Binning) : 不同预估值大小区间为校准的维度



2. 保序回归 (Isotonic Regression) : 利用模型排序能力减小稀疏问题影响



3. 线性插值 (Linear Interpolation) 保证校准函数的平滑性



SIR：保序回归平滑校准算法 (Smoothed Isotonic Regression)

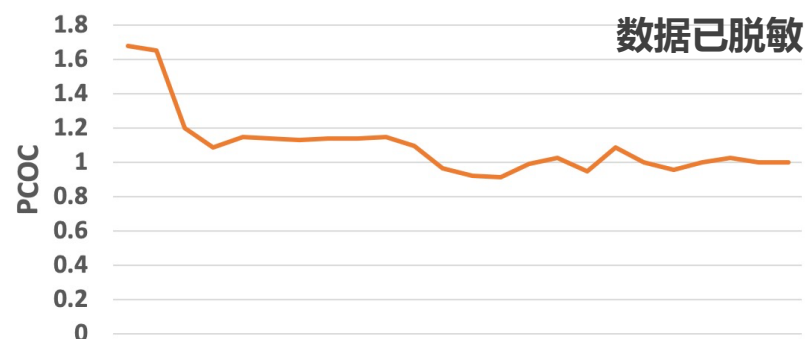
SIR算法在AB两个数据集上的实验对比 (N为分桶数，GC-N表现)

Dataset	Method	$N = 3$	$N = 4$	$N = 5$
A	No calibration	0.63	0.67	0.72
	(1) BBQ	0.57 (-9.5%)	0.63 (-6.0%)	3.05 (+323.6%)
	(2) IR	0.82 (+30.0%)	0.89 (+32.8%)	0.98 (+36.1%)
	(3) Beta calibration	0.56 (-11.1%)	0.60 (-10.4%)	0.66 (-8.3%)
	(4) SIR	0.56 (-11.1%)	0.60 (-10.4%)	0.65 (-9.7%)
B	No calibration	0.40	0.45	0.46
	(1) BBQ	0.71 (+77.5%)	0.81 (+80.0%)	0.84 (+82.6%)
	(2) IR	0.31 (-22.5%)	0.35 (-22.2%)	0.40 (-13.0%)
	(3) Beta calibration	0.27 (-32.5%)	0.32 (-28.9%)	0.36 (-21.7%)
	(4) SIR	0.27 (-32.5%)	0.30 (-33.3%)	0.35 (-23.9%)

Table 3: Business result of A/B test.

Business metrics	RPM	CTR	ROI
Improvement(%)	+3.86%	+8.93%	+5.07%

“计划” 创建后SIR算法的校准效果



预估/校准冷启动问题

新建计划	强时效性计划	小计划 (客户)
• 初始阶段的 投放效果	• 直播	• 公平性 • 活跃度

Bayes-SIR：贝叶斯平滑SIR校准算法

朴素的贝叶斯平滑思想举例：

$$ctr = \frac{click}{pv}$$

先验信息

$$ctr = \frac{click + K_1}{pv + K_2}$$

Eg. $ctr = \frac{click}{pv} = \frac{3}{6} = 50\%$

$K_1 = 20$
 $K_2 = 1000$

$$ctr = \frac{3 + 20}{6 + 1000} = 2.3\%$$

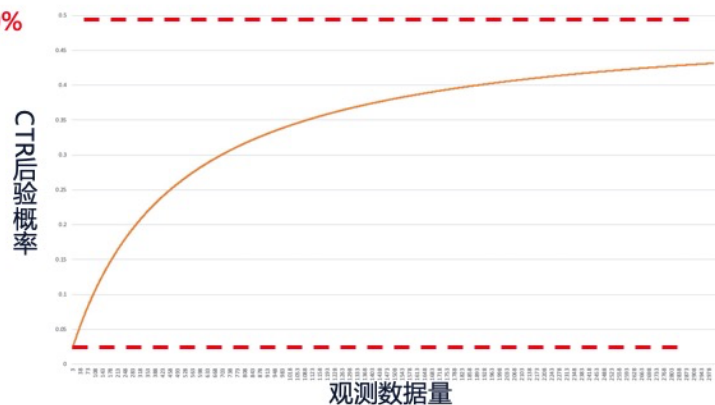
$$ctr = \frac{30 + 20}{60 + 1000} = 4.7\%$$

$$ctr = \frac{300 + 20}{600 + 1000} = 20\%$$

$$ctr = \frac{30000 + 20}{60000 + 1000} = 49\%$$

新观测数据的CTR：50%

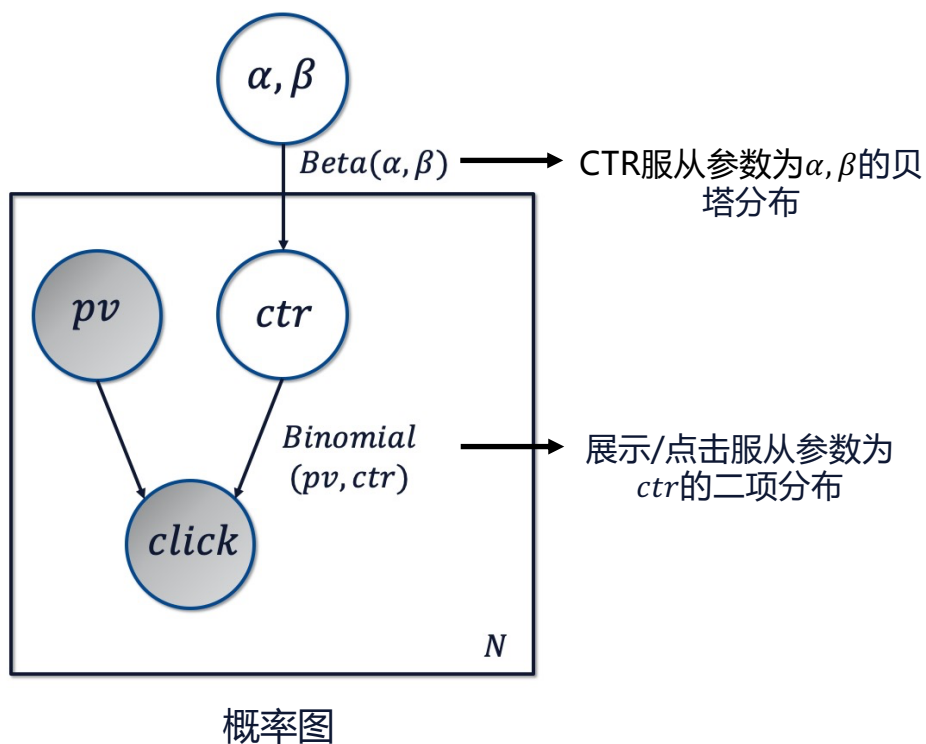
CTR先验概率：2%



新数据积累

Bayes-SIR : 贝叶斯平滑SIR校准算法

应用贝叶斯思想缓解数据稀疏带来的点击率统计不准确问题，结合SIR解决校准冷启动问题。

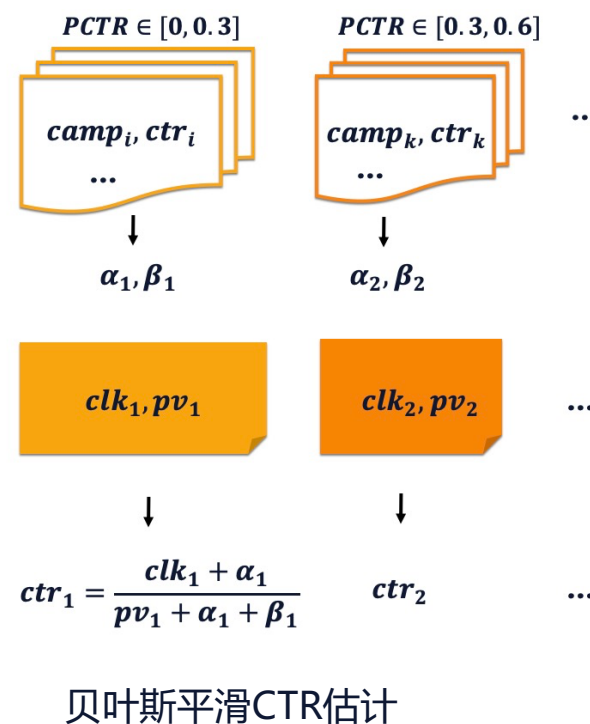


1) 先验数据

2) 贝塔分布
参数求解 (矩估计/似然)

3) 观测数据
(不置信)

4) 后验点击率

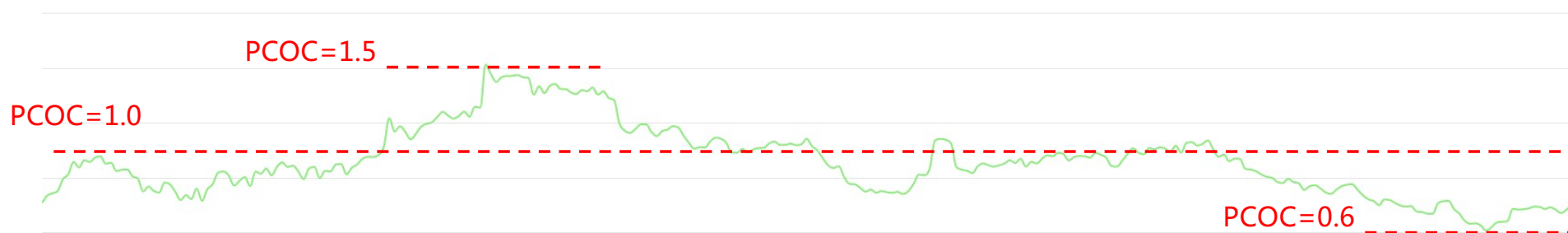


$$P(ctr|X, \alpha, \beta) \propto P(ctr|\alpha, \beta)P(\hat{X} = x|ctr) \propto Be(x + \alpha, n - x + \beta)$$

RTW-BSIR：实时Bayes-SIR校准算法

PCOC的波动问题会严重影响算法出价（oCPX，Auto-Bidding）的效率、稳定性以及客户侧表现。

数据已脱敏

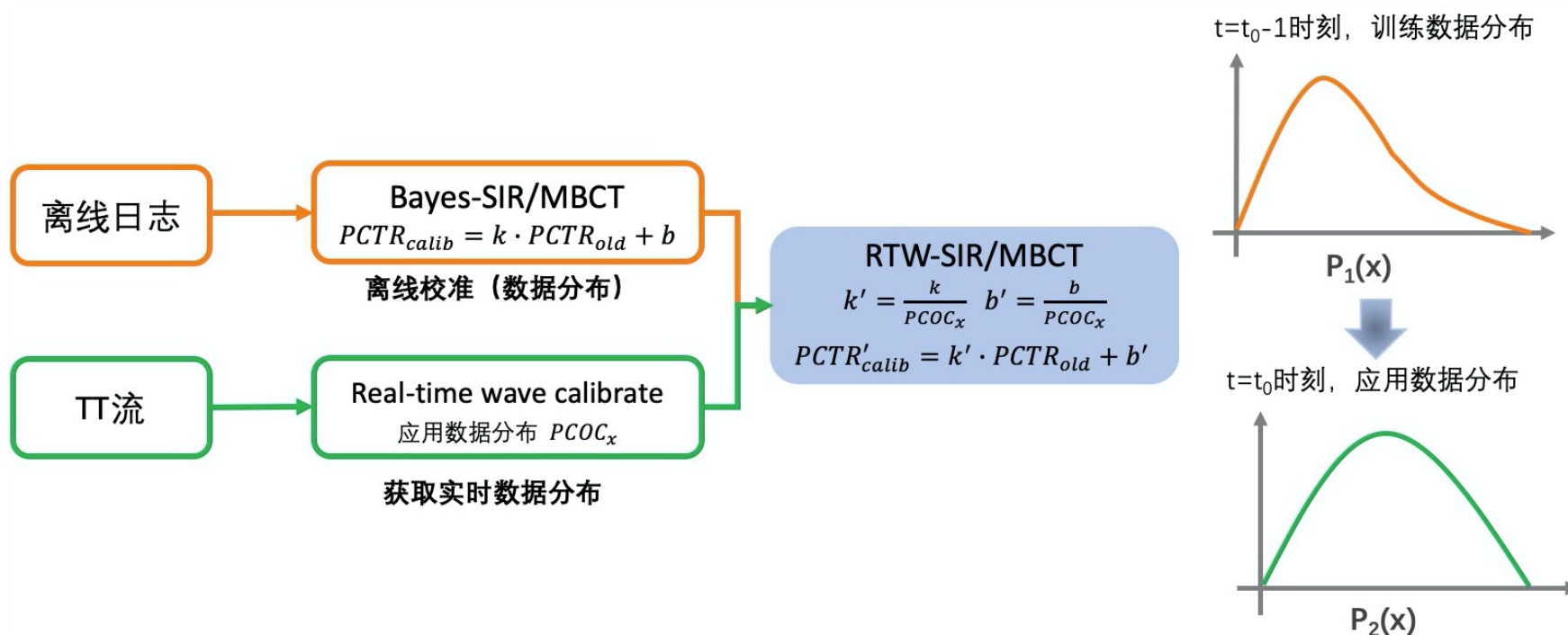


时序上的PCOC波动问题？

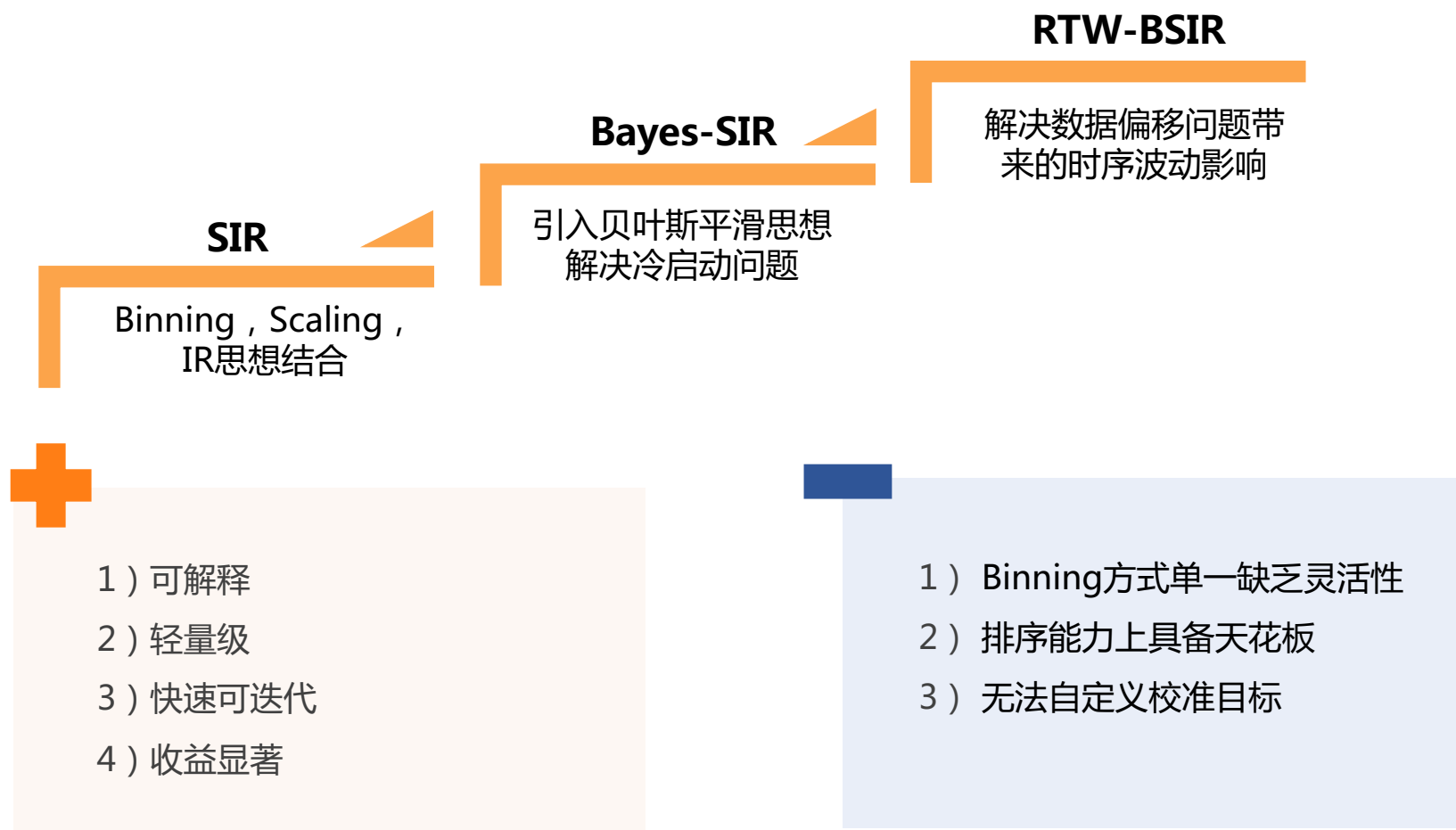
- 1) 加剧环境波动，影响算法策略的稳定性
- 2) 广告主成本（PPC）不可控，影响广告成效和赔付

RTW-BSIR：实时Bayes-SIR校准算法

这是一种典型的**数据偏移 (Dataset Shift)** 问题，解法的核心思想在于：计算训练数据与应用数据的分布差异来修正校准函数（也可采用反事实推断思想进行无偏处理），实现无偏的校准。



SIR系列校准算法小结

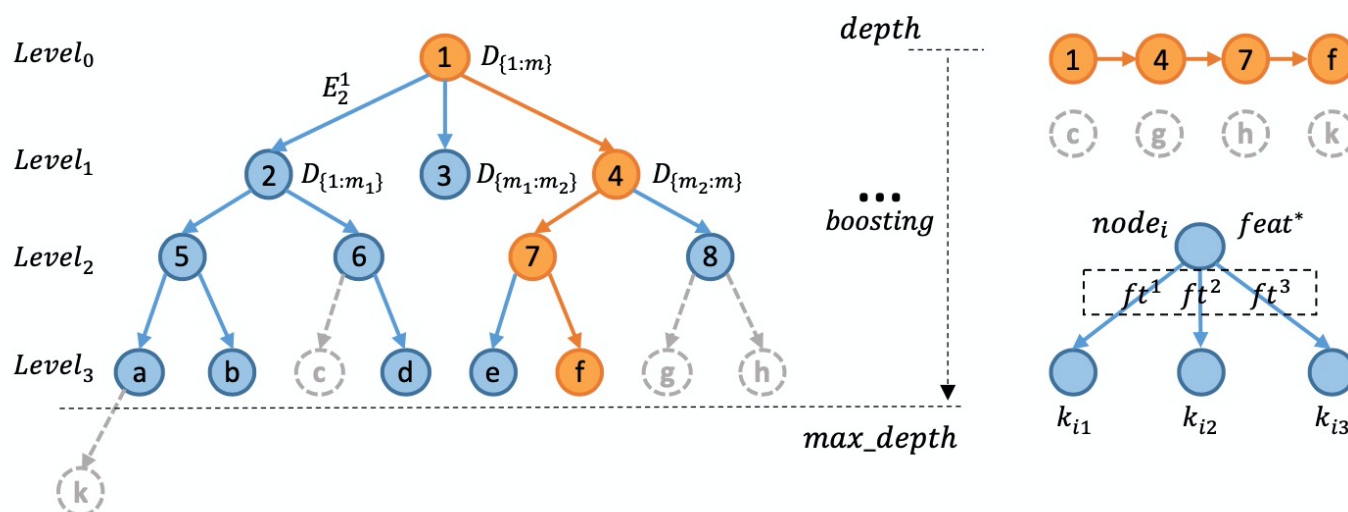


MBCT：基于树结构的Boosting校准技术

Multiple Boosting Calibration Tree

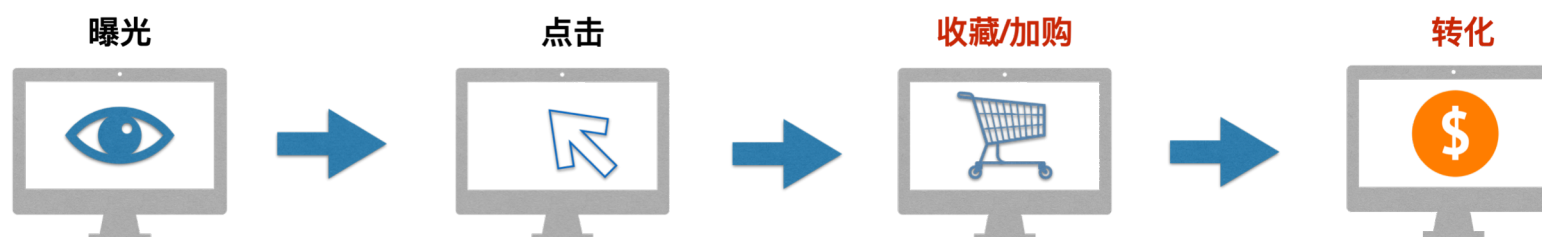
算法思想

- 树结构特性：白盒，可解释，可转化为“IF...THEN”规则
- 节点划分方法：自定义校准目标下启发式学习复杂的分箱规则（Binning）
- Boosting设计：增强模型的表征能力和对稀疏数据的处理能力



后链路预估值校准 (PCCEM)

后链路指标：曝光/点击之后的加购、收藏、转化、关注等指标。



数据稀疏性：后链路指标相比点击更加稀疏，**正样本极少**（如CVR通常小于0.1%）

延迟反馈：后链路指标的反馈延迟通常**有数小时甚至数天**

需求多样性：后链路指标效果优化的**重要性越来越强**，优化指标的**类别越来越多**

后链路预估值校准 (PCCEM)

Post-Click Conversion Estimation Model (PCCEM)

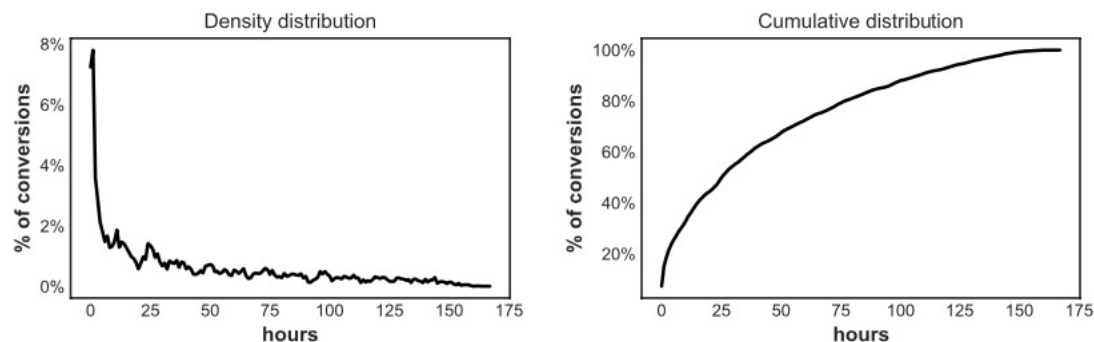


Fig. 3: Distribution of conversion within different time delays.

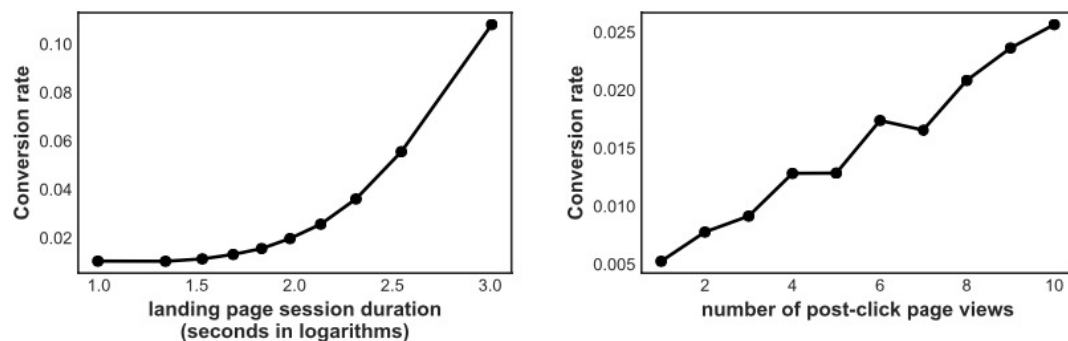


Fig. 4: The relationship between post-click information within one hour after the click and average conversion rate.

Algorithm 2 PCCEM based Calibration for Delayed Response

```

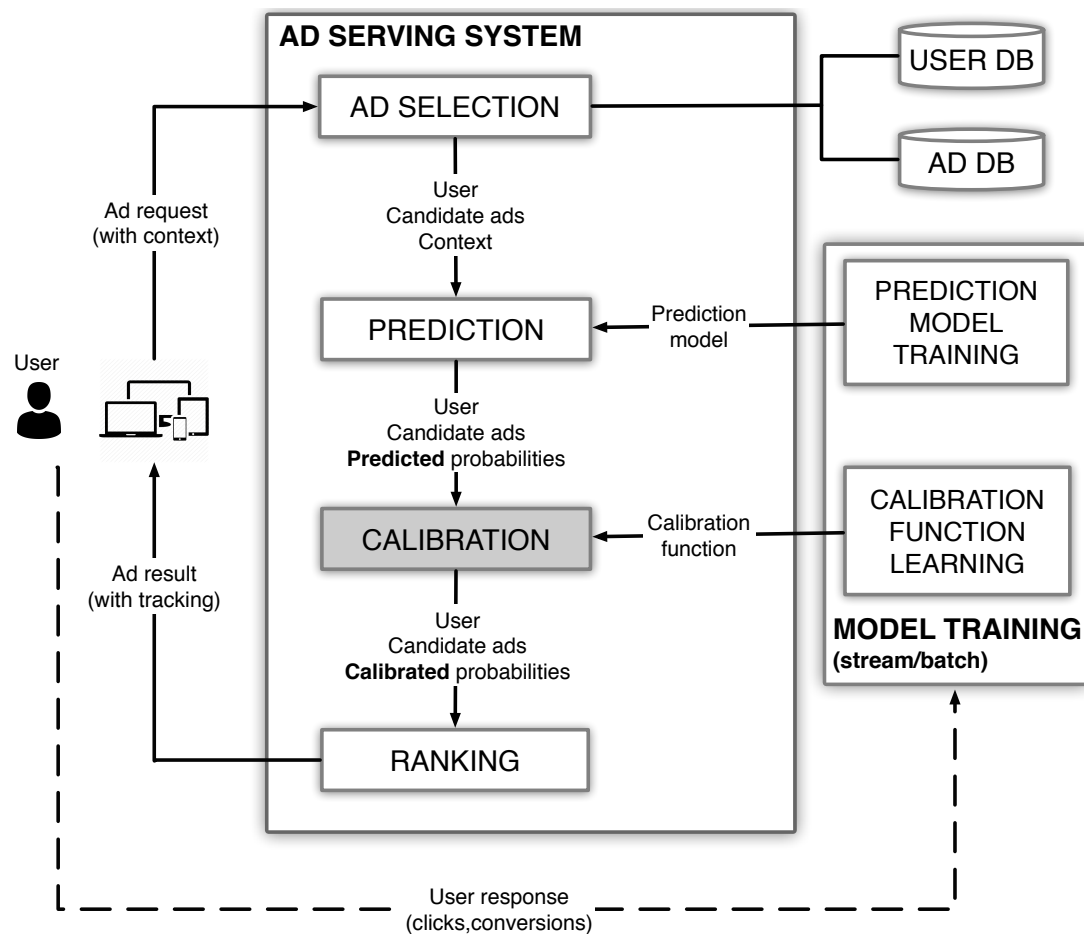
1: for long-term period  $T = 1, 2, 3, \dots$  do
2:   Generate training set  $S$  by using post-click information as features and real
   responses as labels
3:   Fit PCCEM  $m_T$  based on  $S$ 
4:   for each short-term period  $t$  in  $T$  do
5:     for each request  $i$  in time period  $t$  do
6:       Collect post-click information
7:       Use  $m_T$  to produce user response score  $\hat{y}_i$ 
8:       Generate a sample  $(x_i, \hat{y}_i)$ 
9:     end for
10:    Generate sample set  $D$  of recent samples
11:    Update  $f$  by using SIR based on  $D$ 
12:  end for
13: end for
  
```

Table 2: GC- N of three calibration methods for predicted conversion rate calibration.

Method	$N = 3$	$N = 4$	$N = 5$
No calibration	4.46	4.47	4.49
(1) STC	5.67 (+27.1%)	5.91 (+32.2%)	7.11 (+58.4%)
(2) LTC	1.57 (-64.8%)	1.72 (-61.5%)	1.81 (-59.7%)
(3) PCCEM	0.50 (-88.8%)	0.58 (-87.0%)	0.71 (-84.2%)

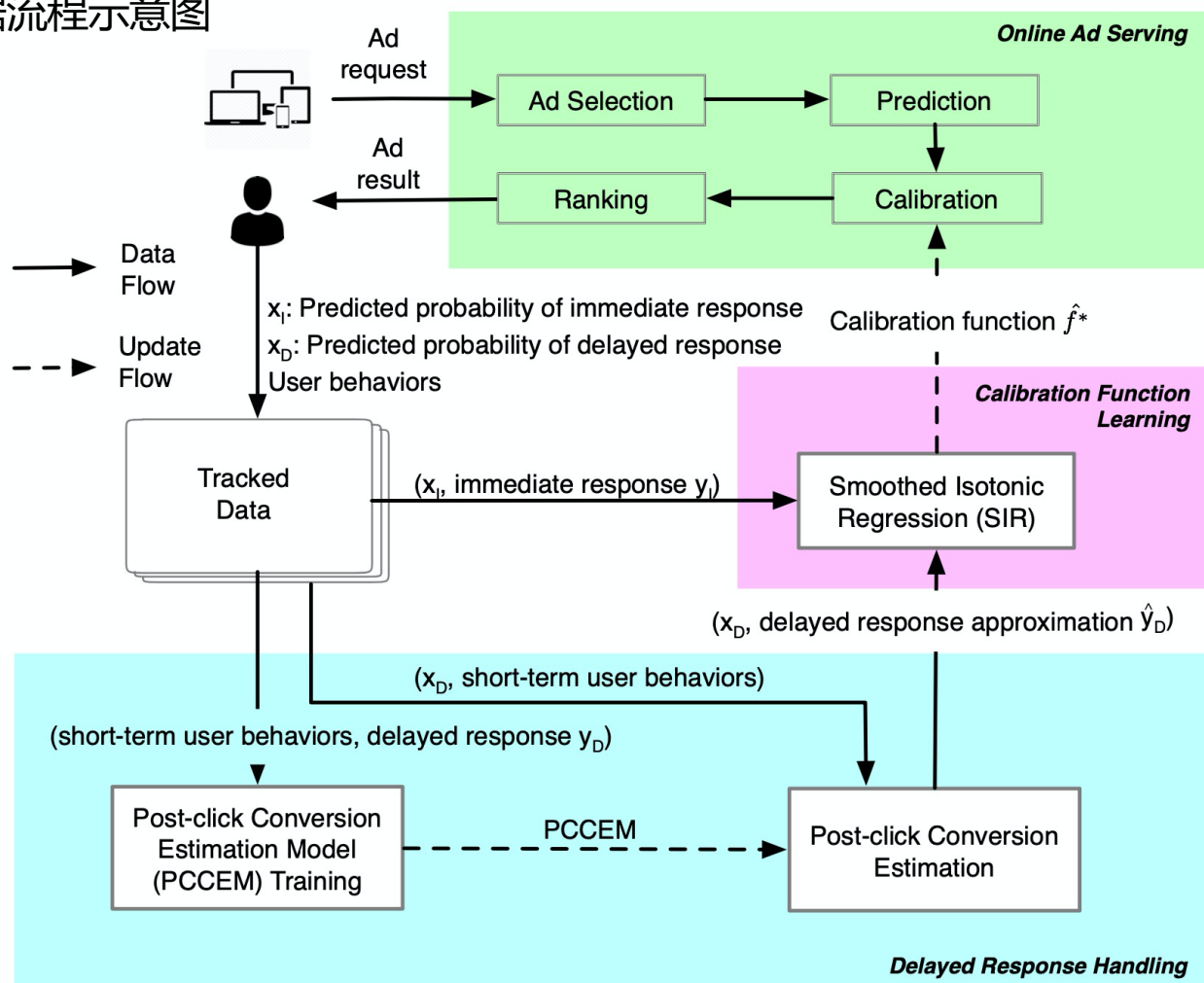
算法落地与工程实践

广告系统模块结构简图



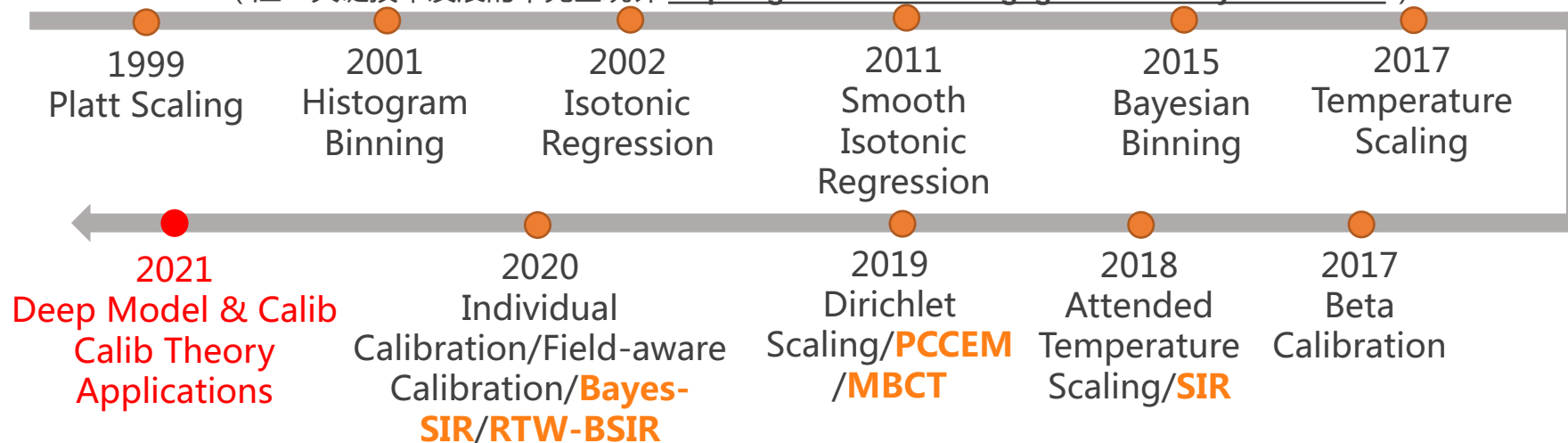
算法落地与工程实践

SIR & PCCEM算法数据流程示意图



总结与展望

(注：关键技术发展的不完全统计 <https://github.com/huangsg1/uncertainty-calibration>)



校准方法的应用场景



附召回定向策略等相关论文（不定期更新）：
<https://github.com/huangsg1/Internet-advertising-mechanism-and-strategy>

Thank You!

我们是阿里妈妈--展示广告算法团队，
欢迎加入合作（siguang.hsg@alibaba-inc.com）！

